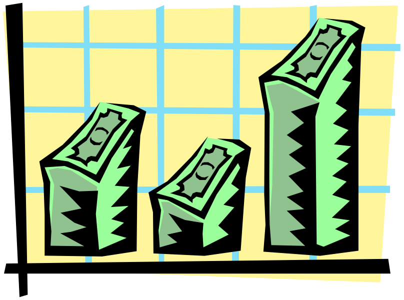
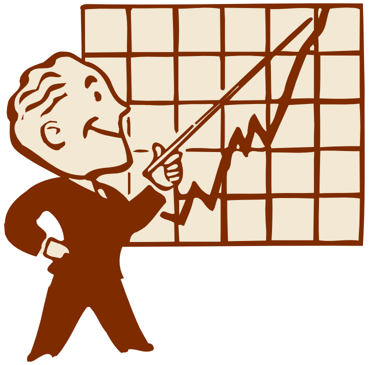




Introduction to Probability and Statistics

Chapter 5

Ammar M. Sarhan,
asarhan@mathstat.dal.ca
Department of Mathematics and Statistics,
Dalhousie University

Fall Semester 2008

Chapter 5

Joint Probability Distributions and Random Samples

Chapter Outlines

5.4 The Distribution of the Sample Mean

5.5 The Distribution of a Linear Combination

Using the Sample Mean

Let $X_1,\ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

$$1.\ E(\overline{X}) = \mu_{\overline{X}} = \mu$$

$$2.\ V(\overline{X}) = \sigma_{\overline{X}}^2 = \sigma^2/n \quad \text{and} \quad \sigma_{\overline{X}} = \sigma/\sqrt{n}$$

In addition, with $T_0 = X_1 + \ldots + X_n$ (The sample total)

$$E(T_0) = n\mu \qquad V(T_0) = n\sigma^2 \qquad \sigma_{T_0} = \sqrt{n}\,\sigma$$

Example 5.24 (p. 213)

Let $X_1,\ldots, X_{25}$ be a random sample from a distribution with mean $\mu=28000$ and standard deviation $\sigma=5000$. Find:

1) $E(\overline{X})$ 2) $V(\overline{X})$

3) E(T) and 4) V(T), where $T = X_1 + \ldots + X_{25}$

Normal Population Distribution

Let $X_1,\ldots, X_n$ be a random sample from a normal distribution with mean value μ and standard deviation σ. Then for any *n:*

1) $\bar{X} \sim N(\mu, \sigma^2 / n)$ 2) $T \sim N(n\mu, n\sigma^2)$

Example 5.25 (p. 214)

The time that it takes a randomly selected rat of a certain subspecies to find its way through a maze is a normally distributed rv with $\mu = 1.5$ min and $\sigma = 0.35$ min. Suppose 5 rats are selected. Find:

1)The probability that the total time for the 5 rats is between 6 and 8 min?

2)The probability that the average time for the 5 rats is at most 2 min?

Solution:

1) $\Phi(0.64) - \Phi(-1.92) = .7115$

2) $\Phi(3.191) = .9993$

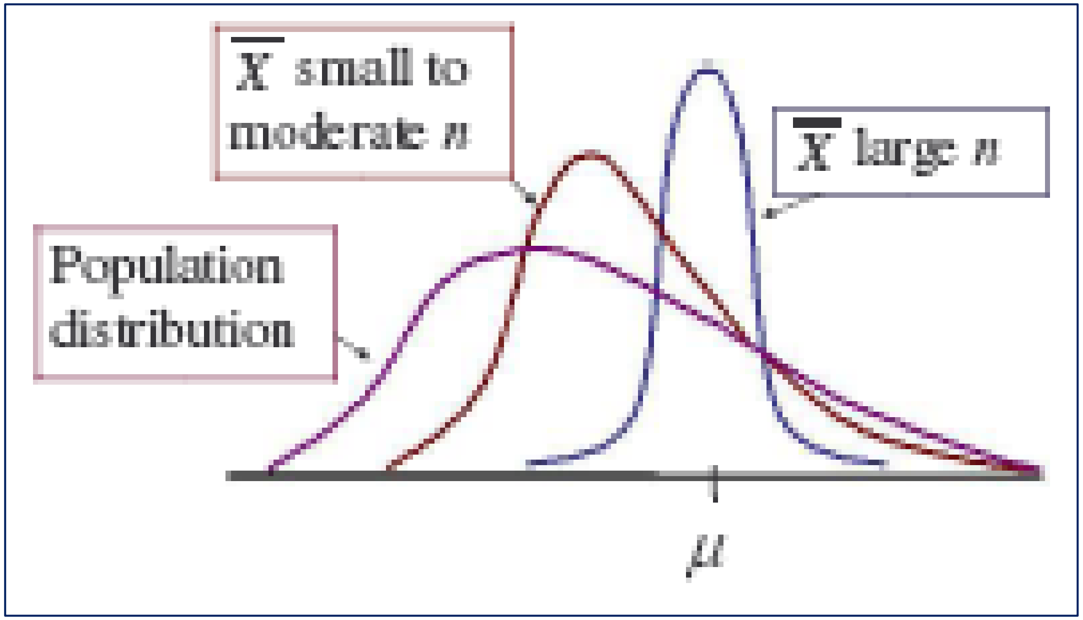
The Central Limit Theorem

Let $X_1,\ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then if n is sufficiently large,

1) $\bar{X} \sim approximately\ N(\mu, \sigma^2 / n)$

2) $T \sim approximately\ N(n\mu, n\sigma^2)$

The larger value of n, the better approximation.



Rule of Thumb

If $n > 30$, the Central Limit Theorem can be used.

Example 5.26 (p. 215)

The amount of a particular impurity in a batch of a certain chemical product is a random variable with mean value 4.0 g and standard deviation 1.5 g. If 50 batches are independently prepared, what is the (approximate) probability that

1)the sample average amount of impurity is between 3.5 and 3.8 g?

2)the sample total amount of impurity is between 175 and 190 g?

Solution:

1) $$n = 50 > 30 \;\Rightarrow\; \bar{X} \sim approximately\, N(4, .2121^2)$$

$$P(3.5 \le \bar{X} \le 3.8) \approx \left(\frac{3.5-4}{.2121} \le Z \le \frac{3.8-4}{.2121} \right)$$

$$= \Phi(\text{-}\ 0.94) - \Phi(\text{-}2.36) = .1645$$

2) $$T \sim approximately\, N(50 \times 4, 50 \times 1.5^2)$$

Linear Combination

Given a collection of *n random variables* $X_1,\ldots, X_n$ *and n numerical constants* $a_1,\ldots, a_n$, the rv

$$Y = a_1 X_1 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

is called a *linear combination* of the X_i's.

Examples:

1) Taking $a_i = 1$, $i=1,2,\ldots,n$, gives $Y = X_1 + \ldots + X_n$ (*Total sum*)

2) Taking $a_i = 1/n$, $i=1,2,\ldots,n$, gives $Y = (X_1 + \ldots + X_n)/n$ (*Sample mean*)

Expected Value of a Linear Combination

Let $X_1, \ldots, X_n$ *have mean values* $\mu_1, \mu_2, \cdots, \mu_n$ and variances of $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$ respectively. Whether or not the X_i's are independent

$$E(a_1 X_1 + \cdots + a_n X_n) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i$$

Variance of a Linear Combination

If $X_1, \ldots, X_n$ are independent

$$V(a_1 X_1 + \cdots + a_n X_n) = \sum_{i=1}^{n} a_i^2 V(X_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

and

$$\sigma_{a_1 X_1 + \cdots + a_n X_n} = \sqrt{a_1^2 \sigma_1^2 + \cdots + a_n^2 \sigma_n^2}$$

Examples:

1) $E(2X_1 - 3X_2) = 2E(X_1) - 3E(X_2)$

2) $V(2X_1 - 3X_2) = 2^2 V(X_1) + (-3)^2 V(X_2)$

3) $\sigma_{2X_1 - 3X_n} = \sqrt{2^2\sigma_1^2 + 3^2\sigma_2^2}$

4) A gas station sells three grades of gasoline: regular, extra, and super. These are priced at \$21.2, \$21.35, and \$21.5 per gallon, respectively. Let X_1, X_2 and X_3 denote the amount of these grades purchased (gallon) on a particular day. Suppose X_i's independent with $\mu_1=1000$, $\mu_2=500$, $\mu_3=300$ and $\sigma_1=100$, $\sigma_2=80$, $\sigma_3=50$. Find

a) The expected revenue from sales on that day;

b) The variance of the revenue from sales;

c) The standard deviation of the revenue from sales;

Solution:

Let Y be the revenue from sales. Then $Y = 21.2X_1 + 21.35X_2 + 21.5X_3$

a) $E(Y) = \$4,125$ b) $V(Y) = 104,025$ c) $\sigma_Y = \$322.529$

Variance of a Linear Combination

For any $X_1, \ldots, X_n$,

$$V\left(a_1 X_1 + \cdots + a_n X_n\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \, Cov\left(X_i, X_j\right)$$

where

$$Cov\left(X_i, X_j\right) = E\left[\left(X_i - \mu_i\right)\left(X_j - \mu_j\right)\right]$$

Difference Between Two Random Variables

$$E(X_1 - X_2) = E(X_1) - E(X_2)$$

and, if X_1 and X_2 are independent

$$V(X_1 - X_2) = V(X_1) + V(X_2)$$

Example: A certain automobile manufacture equips a particular model with either a six-cylinder engine or a four-cylinder engine. Let X_1 *and* X_2 be fuel efficiencies for independently and randomly selected six-cylinder and four-cylinder cars, respectively. With $\mu_1=22$, $\mu_2=26$, $\sigma_1=1.2$ and $\sigma_2=1.5$. Find

a) $E(X_1 - X_2)$

b) $V(X_1 - X_2)$

c) $\sigma_{X1 - X2}$

Difference Between Normal Random Variables

If $X_1, X_2, \ldots X_n$ are independent, *normally* distributed rv's, then any linear combination of the X_i's also has a normal distribution. The difference $X_1 - X_2$ between two independent, normally distributed variables is itself normally distributed.

Example: In example 4, slide 10, if X_i's are normally distributed find:
(1) probability that revenue exceeds 4500?
(2) Probability that the amount of regular gas purchased exceeds that of supper by 500 gallon?

Solution:

(1) $Y \sim N(4{,}125,\ 104{,}025)$, then $P(Y > 4500) = P\left(Z > \dfrac{4500 - 4125}{322.53}\right)$

$$= P(Z > 1.16) = 1 - \Phi(1.16) = 0.123$$

(2) In class.